Algorithm Theoretical Basis Document

D2 The Next-Generation Data Assimilation System

Data Assimilation Office
Goddard Space Flight Center
Greenbelt, Maryland 20771
January 2000

Contents

1 Introduction

The purpose of this document is to define the scientific attributes of DAO's next-generation *Core* Data Assimilation System (Core DAS) currently being developed at NASA's Data Assimilation Office. More specifically, we focus on the first configuration of a Core DAS based on the NASA-NCAR General Circulation Model and the Physical-space Statistical Analysis System; we will refer to this early configuration as the *Violet* Core system. This document will concentrate on establishing the scientific requirements for the *Violet* system; software issues are not addressed here. In particular, issues related to I/O, pre- and post-processing, as well as monitoring, are beyond the scope of this document.

The *Violet* Core System consists of the following main components:

General Circulation Model. The General Circulation Model used for this Core DAS is the model jointly developed by the Data Assimilation Office (DAO) and the Climate and Global Dynamics Division (CGDD) at NCAR. This model is based on the *finite-volume dynamical core* developed at DAO (Lin and Rood 1996, Lin and Rood 1997, Lin 1997, Lin and Rood 1998) with physical parameterizations from the NCAR CCM3 (Kiehl et al. 1996). For additional details see the companion document *DAO ATBD / Next-Generation Model*.

Quality Control. The Statistical Quality Control (SQC) System is used to screen observational data prior to assimilation. This QC system consists of simple check of the observations against a background field, followed by an adaptive buddy check which adjusts error bounds according to the *flow of the day*.

Analysis System. The Physical-space Statistical Analysis System (PSAS, Cohn et al 1998) is used to combine a first guess from the NASA-NCAR GCM with observational data to provide an updated estimate of the state of the atmosphere.

The NASA-NCAR GCM is a completely new model which which replaces the GEOS GCM used in the GEOS-1/2/3 Data Assimilation systems; see DAO (1996) and the companion document *DAO ATBD / GEOS-Terra* for description of GEOS GCM. A particular configuration of SQC and PSAS are currently implemented in GEOS-Terra system which is about to become operational. However, the unique finite-volume formulation of the NASA-NCAR GCM, combined with the generality of the observation-space formulation of PSAS, call for a complete redesign of the current GEOS DAS model-analysis interface which has its roots in the *Optimum Interpolation* (OI) algorithm of GEOS-1 DAS (Pfaendtner et al. 1994). The main characteristics of this new model-analysis interface are:

- The system works in both *observation* and *finite-volume* spaces, with no intermediate constant pressure transformations necessary. Therefore, the *p-to-σ* and *σ-to-p* interpolations of GEOS DAS have been eliminated.
- Observations are accurately simulated from model fields by mapping volume-mean fields into observables. Proper consideration is taken of the nature of the observable (*e.g.*, layer-mean vs. point measurements).
- PSAS now produces analysis increments directly on model grid, this way preserving the balance relationships implied by the error covariance formulation.

- The surface and upper-air analyzes have been unified, thus ensuring the consistency between surface pressure and low-level geopotential height analysis, and maximizing the impact of surface wind observations on the upper-air fields.

- The initial *Violet* Core System will include a Rapid Update Cycle (RUC) with analysis performed approximately every hour.

- As shown theoretically in Bloom et al (1995), and confirmed by ongoing RUC research based on GEOS DAS (da Silva and Lou, personal communication), the incremental analysis update (IAU) scheme of GEOS DAS is quite ineffective as a low-pass filter when the analysis interval is shortened to one hour. Therefore, the *Violet* Core System will not include IAU. The need for IAU or any other form of *initialization* during the assimilation cycle will be determined by our experience with this initial system.

- In part due to the elimination of IAU, the extended (5-day) forecast procedure will include a simple digital filter aimed at improving forecast skills.

This document is organized as follows. A general description of the sequential data assimilation algorithm is presented in section 2, identifying the main system components. The NASA-NCAR general circulation model is described in companion document *DAO ATBD / Next-Generation Model*. The general idea behind the adaptive Statistical Quality Control (SQC) System is introduced in section 3, along with a description of the main observation screening tests. The Physical-space statistical analysis system is outlined in section 4, with a comprehensive list of PSAS technical references. Besides the new model, the second most unique feature of this data assimilation system is how we interface the finite-volume GCM fields to the physical-space formulation of PSAS; the details of this interface appear in section 5. Some of the issues relating to the physical balance of analysis state and the potential need for initialization are discussed in section 6. The initial system configuration, including resolution, and the overall validation strategy are given in section 7. A partial list of open problems and issues which have not been addressed by this initial *Violet* Core system are included in section 8. Concluding remarks appear in section 9.

2 Algorithm overview

2.1 Notation

n	number of gridpoints $\times$ variables	$n \sim 10^6$
p	number of observations	$p \sim 10^5$
w^a	gridded analysis state vector	$\in I\!R^n$
w^f	gridded forecast state vector	$\in I\!R^n$
w^o	observation vector	$\in I\!R^p$
$\mathcal{I}$	interpolation operator	$\mathcal{I} : I\!R^n \rightarrow I\!R^p$
f	non-linear observation operator	$f : I\!R^p \rightarrow I\!R^p$
F	tangent linear version of f, $F = \partial f / \partial w$	$F : I\!R^p \rightarrow I\!R^p$
h	generalized interpolation operator, $h(w) = f(\mathcal{I}w)$	$h : I\!R^p \rightarrow I\!R^n$
H	tangent linear version of h, $H = F\mathcal{I}$,	$H : I\!R^p \rightarrow I\!R^n$
a	non-linear forecast model	$a : I\!R^n \rightarrow I\!R^n$
A	tangent linear version of a	$A : I\!R^n \rightarrow I\!R^n$
P^f	forecast error covariance	$P^f : I\!R^n \rightarrow I\!R^n$
R	observation error covariance	$R : I\!R^p \rightarrow I\!R^p$
P^a	analysis error covariance	$P^a : I\!R^n \rightarrow I\!R^n$
Q	model error covariance	$Q : I\!R^n \rightarrow I\!R^n$

2.2 The Sequential Data Assimilation Algorithm

In this section we briefly outline the *Extended Kalman Filter* algorithm (*e.g.*, Cohn 1997 and references within) which provides the theoretical framework for the operational data assimilation system that we implement. Suppose we are given a non-linear forecast model

$$w_k^f = a_{k-1}(w_{k-1}^a) \tag{1}$$

where w_k^f is the forecast vector at time t_k and w_{k-1}^a is the analysis vector at the previous synoptic time. Having a gridded forecast state, one proceeds to compute *innovations* or observation minus forecast residuals. The forecast state variables are related to observations through a generalized interpolation operator h, viz.

$$v_k = w_k^o - h(w_k^f) = w_k^o - f(\mathcal{I}w_k^f) \tag{2}$$

The *observation operator* f acts on state variables at observation locations; the operator $\mathcal{I}$ interpolates state variables from grid points to observation locations. Neglecting forecast error bias for the moment (Dee and da Silva 1998), the sequential EKF algorithm is given by the following set of equations:

$$\begin{aligned}
w_k^f &= a_{k-1}(w_{k-1}^a) && (3)\\
P_k^f &= A_{k-1} P_{k-1}^a A_{k-1}^T + Q_{k-1} && (4)\\
P_k^a &= (I - K_k H_k)\, P_k^f && (5)\\
y_k &= \left[F_k \mathcal{I}_k P_k^f \mathcal{I}_k^T F_k^T + R_k \right]^{-1} \left(w_k^o - f(w_k^f) \right) && (6)\\
w_k^a &= w_k^f + P_k^f \mathcal{I}_k^T F_k^T y_k && (7)
\end{aligned}$$

In the *Violet* Core System, forecast step in (3) will be accomplished by integrating the NASA-NCAR atmospheric general circulation model, while (6)—(7) are solved with the Physical-space Statistical Analysis System (PSAS, Cohn *et al.* 1998). The forecast error covariance equation (4), as well as the analysis error covariance equation (5) are computationally too expensive for a *brute force* implementation. Although approximate algorithms for the evolution of the calculation of these matrices are being developed (*e.g.*, Cohn and Todling 1996, Todling et al 1999), current operational practice is to specify these matrices from simple statistical models. A brief description of the error covariance modeling in GEOS-2 DAS is described in DAO's *Algorithm Theoretical Basis Document* (DAO 1996). Ongoing research on the *Parameterized Kalman Filter* (PKF) at DAO is aimed at developing practical approximations to (4)-(5) which incorporate the important effects of flow dependency and anisotropy absent in the current operational implementation of PSAS. This algorithm refinement is beyond the scope of the *Violet* Core System.

3 Statistical Quality Control (SQC)

The on-line Statistical Quality Control (SQC) system attempts to identify observations that are likely to be contaminated by gross errors. The algorithms involve statistical tests of the actual data against assumptions about their expected errors and about GCM forecast errors. Essentially, a local statistical analysis is performed for each outlier observation, i.e., for each observation that differs significantly from the short-term forecast produced by the GCM. If this analysis indicates that the observation is inconsistent with surrounding data, then that observation is marked for rejection.

The SQC encompasses a *background check*, a *buddy check*, a *wind check*, and a *profile check*, each of which is described below. All checks are formulated in terms of the observed-minus-forecast residuals (O-F) rather than the observations themselves. All checks potentially modify the quality control marks associated with the observations, but leave all other data attributes unchanged. The background check and buddy check involve the forecast and observation error variances for the quantities being tested, which are prescribed in the global analysis system.

3.1 Statistical aspects

The SQC algorithms operate on the vector of observed-minus-forecast residuals v defined by

$$v = w^o - f(\mathcal{I}w^f), \tag{8}$$

where w^o is the vector of observations, w^f is the forecast vector, f is the observation operator, and $\mathcal{I}$ is the linear operator which interpolates state variables from model grid points to observation locations. The observation operator maps model variables to observables. For remotely sensed radiances, for example, the function f represents a radiative transfer model. It is simply the identity for conventional, *in situ* observations of model variables.

The SQC attempts to identify corrupt data based on statistical expectations. This requires knowledge of the covariance S of the observed-minus-forecast residuals, defined

by

$$S_{ij} = \langle v_i v_j \rangle, \tag{9}$$

with i, j indicating location. In general these covariances are poorly known, but a rough estimate is available from the global analysis system. It follows from (8) that

$$S \approx F\mathcal{I}P^f\mathcal{I}^T F^T + R, \tag{10}$$

where F is the linearized observation operator

$$F = \left.\frac{\partial f}{\partial w}\right|_{w=w^f}, \tag{11}$$

and P^f, R are the covariances of forecast and observation errors, respectively. Equation (10) would be exact if forecast and observation errors were entirely independent (they are not, since both types of errors depend on the true state) and if all observation operators were linear.

Specification of reasonably accurate error covariances is crucial to the quality of a statistical analysis. We therefore assume that the right-hand side of (10), as prescribed by the global analysis system, provides some useful information about the residual error covariances. Accordingly, prescribed error statistics are used to define tolerances for the background check, whose main purpose is to mark outlier observations for subsequent reexamination in the buddy check. However, since actual errors depend on many unknown model defects and other intangibles, covariance specifications in operational data assimilation systems cannot be relied upon to accurately describe error characteristics in all situations at all times. In particular, during extreme events—when quality control decisions become especially important—the covariances as prescribed by the global analysis system are almost certainly inadequate. Thus, a key aspect of the SQC is the attempt to adjust the prescribed error statistics based on actual data. This adjustment takes place during the buddy check, before a final accept/reject decision is reached for an outlier observation.

3.2 The background check

The background check tests each single observation against a background estimate, which is simply the 6-hour model forecast interpolated to the time and location of the observation. If the discrepancy is extremely large then the observation is rejected outright. If the discrepancy is large, but not extremely large, then the observation is marked as an outlier, to be reexamined in the buddy check. The tolerances for the background check are defined in terms of standard deviations obtained from the error statistics as prescribed by the global analysis system.

The algorithm is as follows:

For each observation w_i^o:

mark w_i^o as an **outlier** if $|v_i| > \tau_o \sigma_i$
mark w_i^o as **excluded** if $|v_i| > \tau_x \sigma_i$

Here $\sigma_i = \sqrt{S_{ii}}$, and τ_s, τ_x are prescribed non-dimensional tolerance parameters. Typically we take $\tau_o = 2, \tau_x = 10$.

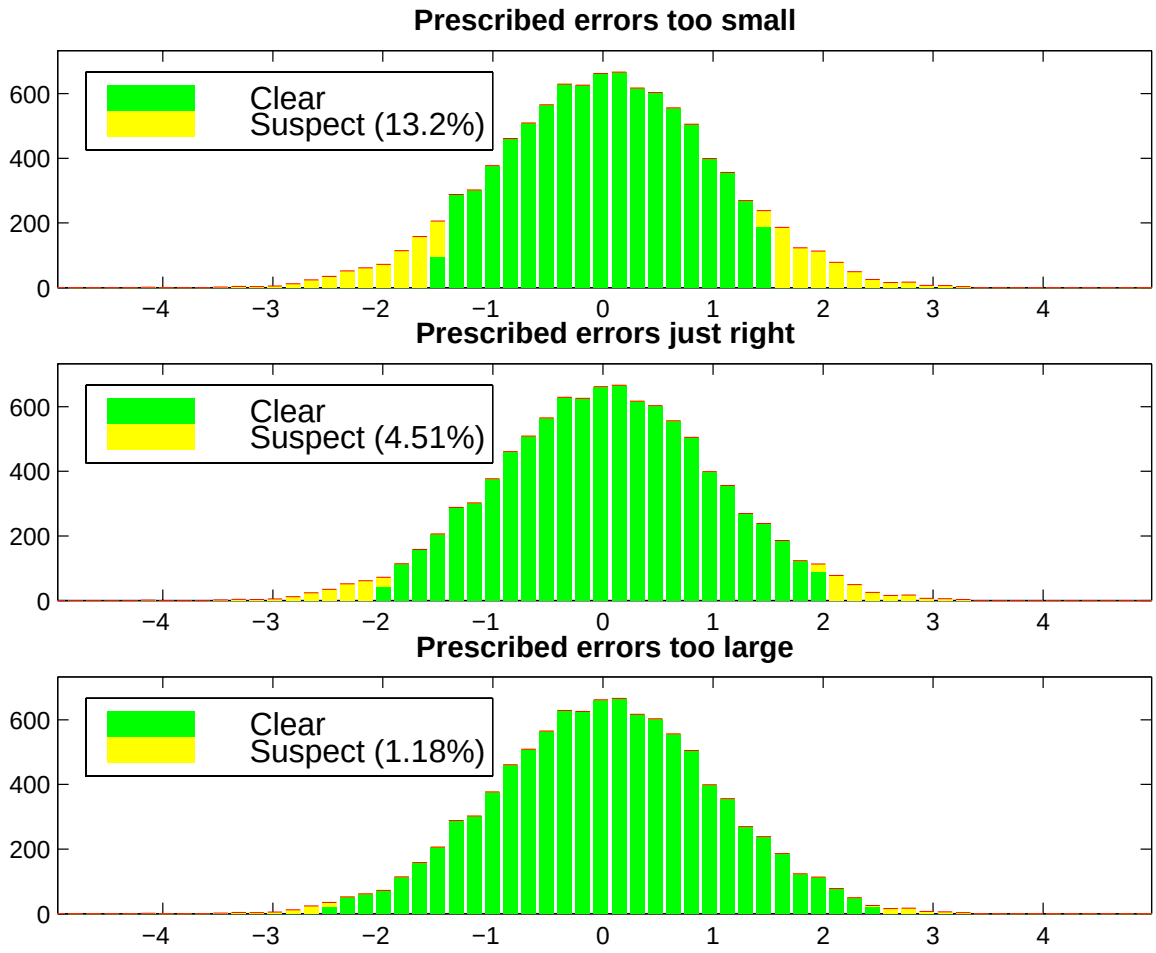

Figure 1: Illustration of the relationship between the rate at which the background check marks observations as outliers and the prescribed error statistics, for normally distributed errors. The yellow tails of the histograms correspond to observations marked as outliers.

The rate at which the background check produces suspect marks presents a useful check on the accuracy of the prescribed error statistics. If the forecast and observation error variances are correctly tuned, and if the errors are roughly normally distributed, then the suspect rate can be predicted. For example, when $\tau_o = 2$, the rate should be about 4.5%. If the actual suspect rate is larger (smaller), then the prescribed error variance is too small (large). This is illustrated in Fig. 1. Monitoring the background check failure rates for specific instruments has, in a number of cases, led to adjustments of observation error statistics in GEOS DAS.

3.3 The buddy check

The buddy check is applied to a subset of observations which are considered suspect, either because they were identified as an outlier by the background check, or because they were marked as suspect during the preprocessing stage. The buddy check attempts to predict the value of a suspect observation from nearby non-suspect

observations (the *buddies*.) If the predicted value is in reasonable agreement with the observation, then the observation is no longer considered suspect. If a sufficient number of buddies is available, then the tolerance for the buddy check is adjusted based on a local estimate of O-F standard deviations. Once all suspect observations have been tested, the entire process is repeated for all observations that are still considered suspect. The process stops when the set of suspects no longer changes: all remaining suspects are then rejected.

The buddy check initially labels observations as suspect based on their quality control history. A single iteration of the algorithm is as follows:

For each suspect observation w_j^o:

1. Define the set of buddies:

   Nearby non-suspect observations of the same data type as w_j^o are ranked according to the scalar weight that each would receive in an optimal univariate statistical analysis at the location of w_j^o. The buddies are simply the n highest ranking of these, where n is a configuration parameter. Typically we take $n = 50$.

2. Predict the value of the suspect observation based on its buddies:

   Using the weights determined in the previous step, the weighted average $v_j^\star$ of the v_i associated with the buddies provides the optimal univariate analysis of the buddies at the location of w_j^o.

3. Adjust the prescribed estimate of the local O-F standard deviation:

   If $\widehat{\sigma}_j^2$ is the sample variance of the v_i associated with the buddies, the prescribed variance σ_j^2 is adjusted according to

   $$(\sigma_j^\star)^2 = (n^\star \sigma_j^2 + n\widehat{\sigma}_j^2)/(n^\star + n) \tag{12}$$

   where $n^\star$ is a configuration parameter. Typically we take $n^\star = 25$.

4. Reevaluate the status of w_j^o:

   Change the status of w_j^o to non-suspect if

   $$|v_j - v_j^\star| < \tau_b \sigma_j^\star \tag{13}$$

   where τ_b is a prescribed non-dimensional tolerance parameter. Typically we take $\tau_b = 3$.

These steps are repeated until no further observations change status. At that point, any remaining suspect observations are marked for rejection.

The adaptive nature of the buddy check has two important consequences. First, the final quality control decisions are not very sensitive to the prescribed error statistics in the global analysis system. We have verified this experimentally by varying the tolerance parameter τ_o of the background check. It was found that the final accept/reject status of observations is not very sensitive to the background check failure rate, as long as this rate is roughly between 1% and 10%. This insensitivity to the prescribed

statistics is a major practical advantage, since (1) these statistics are not very reliable and (2) the SQC algorithms do not require retuning each time the prescribed statistics in the global analysis change.

The second consequence of adjusting rejection limits on the fly based on the local variability of surrounding data is that the buddy check becomes increasingly tolerant in synoptically active situations (and, conversely, more stringent when the flow is smooth). This is best illustrated by an example, in which we contrast the results of a nonadaptive buddy check against those of the adaptive buddy check. Figure 2 shows two maps with quality control marks for zonal wind observations (obtained from aircraft and rawinsonde reports) over North America at or near 200hPa, on January 14 1998. The top panel shows rejections (indicated by red marks) by a non-adaptive buddy check, based on tolerances derived from prescribed statistics. Yellow marks indicate data that were marked as outliers by the background check, but which passed the buddy check. The lower panel shows rejections by the adaptive buddy check. Tolerances are increased due to greater variability than implied by the prescribed statistics, resulting in the acceptance of several additional outlier observations. The effect on the wind analysis (not shown) is to increase wind speeds by about $3m/s$ in some places.

3.4 The wind check

This check is applied to all u-wind and v-wind data to make sure that wind components pass the quality control in pairs. The algorithm determines whether two wind components are paired (i.e., whether they originate from the same report) by matching their location attributes, instrument type, and sounding index.

3.5 The profile check

This check eliminates an entire vertical sounding in case any of the data from that sounding are marked for exclusion. It is applied to selected data types only. Currently the profile check is used for TOVS height retrievals only. For example, if the buddy check rejects a TOVS height observation at 10hPa, then the entire sounding is marked for rejection.

3.6 Special treatment of moisture observations

The analysed moisture field in GEOS DAS is water vapor mixing ratio, which is highly variable in space and time. This causes difficulties for the buddy check, which presumes that the field is spatially coherent on the scales resolved by the observing network. Experience has shown that a buddy check applied to water vapor mixing ratio observations (or, equivalently, specific humidity) tends to reject too many of them, unless the tolerances are relaxed to a point where the quality control becomes almost completely inactive. This is obviously not acceptable, unless preprocessing quality control is completely reliable.

To remedy this situation, the statistical tests (background check and buddy check) in the SQC are applied to relative humidity residuals. These residuals are computed in

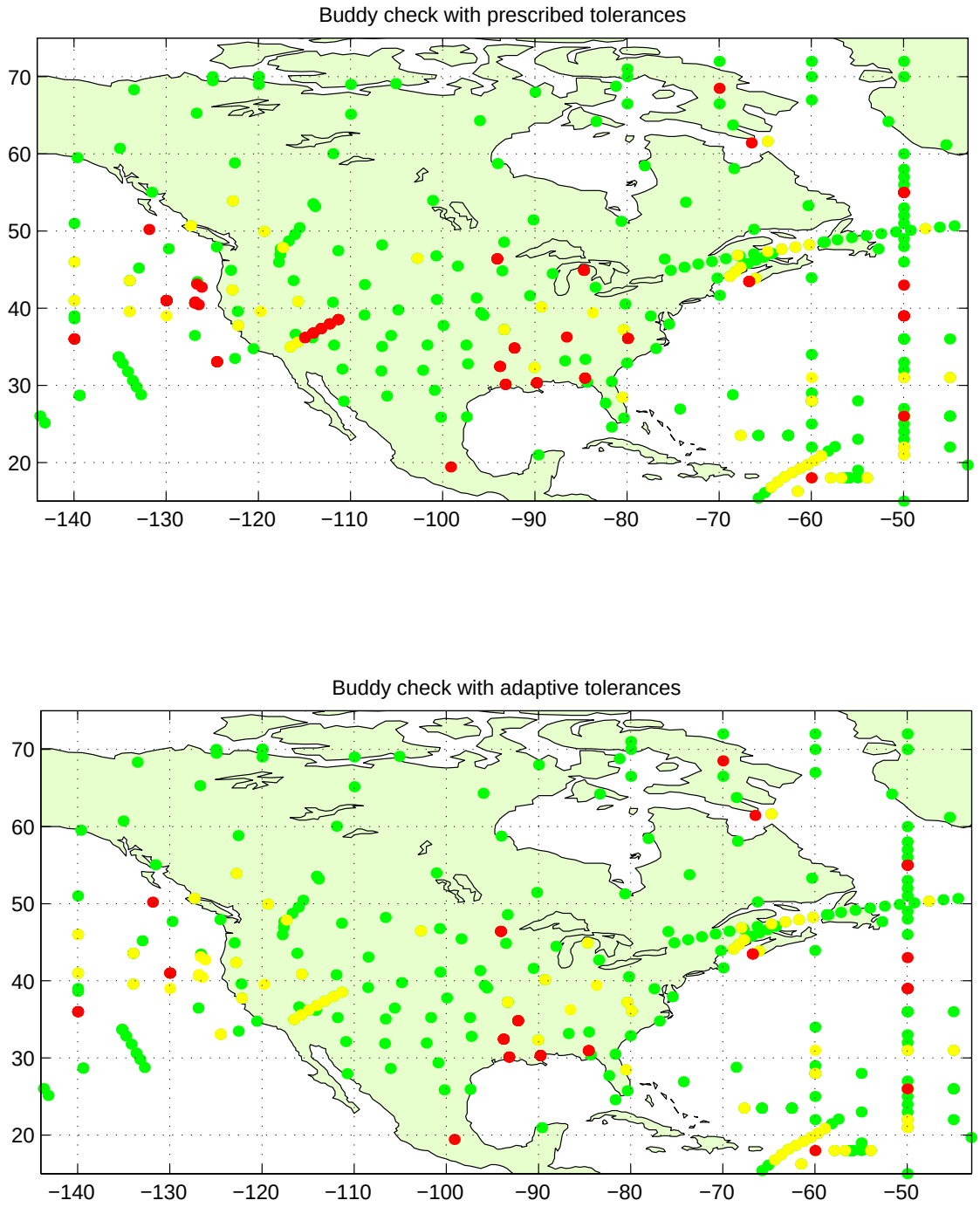

Figure 2: Quality control decisions for zonal wind observations at 200hPa on January 14 1998, using a non-adaptive buddy check (top) and adaptive buddy check (bottom).

two ways: first, using observed mixing ratios and observed temperatures, and second, using observed mixing ratios and model-predicted temperatures. This prevent the situation in which a relative humidity looks good even though both mixing ratio and temperature are corrupt. The tests are applied in sequence to both types of residuals, and an observation passes QC only if none of the tests fail.

4 The Physical-space Statistical Analysis System (PSAS)

4.1 The PSAS Solver

The PSAS algorithm solves the analysis equations (6) and (7) in a straightforward manner. The *innovation covariance matrix*

$$M \equiv HP^f H^T + R \tag{14}$$

entering (6) is symmetric positive definite, making a standard pre-conditioned conjugate gradient (CG) algorithm (Golub and van Loan 1989) the method of choice for solving the large linear system (6). For the current observing system, setting up and solving the linear system (6) represents about half the computational effort of PSAS, and involves computation in observation space: $M \in I\!R^{p\times p}$ and $y \in I\!R^p$. The other half of the computational expense is taken by step (7) which transfers the solution y to the state space: $P^f H^T y \in I\!R^n$. For additional technical details on the implementation of the PSAS solver consult da Silva and Guo (1996), Guo and da Silva (1997), Guo et al (1998), Larson et al (1998)

4.2 Error covariance modeling

The error covariance models used for the initial *Violet* Core system will be the same implemented for the GEOS-Terra system, with model parameters retuned as appropriate. These models were originally described in the ATBD (DAO 1996), with a more up to date version appearing in da Silva et al (1999). The main characteristics of these models are:

- PSAS employs compactly supported spline functions for modeling all single-level univariate correlations. This means that the modeled horizontal correlations are exactly zero beyond a certain finite distance (usually 6000 km). Currently, these univariate correlation are horizontally isotropic. Three-dimensional covariances are constructed in terms of single-level isotropic covariances.
- Geopotential height and mixing ratio errors are assumed uncorrelated.
- Wind-mass covariances are modeled according to a linear friction balance which ensures the geostrophic balance of the analysis increments in the extra-tropics, and cross-isobar flow near the surface and in the tropics.

- In addition to the wind-mass balanced component, wind errors are allowed to possess a unbalanced component which permits height decoupled wind analysis increments in the tropics.
- Error covariance parameters entering these models are estimated with the maximum-likelihood procedure of Dee and da Silva (1999).
- Height errors variances are adaptively estimated from observation minus forecast residuals from radiosondes and TOVS retrievals. These variances can capture the variability of the height error variances on time scales longer than 15 days.

The initial configuration of the *Violet* Core system will rely on error covariance parameters tuned for the GEOS-2 DAS system. Once the system has been run for a period of a month these parameters will be re-tuned using the output ODS files (da Silva and Redder 1995). The forecast error variances will also be reestimated using the adaptive algorithm described in DAO (1996).

4.3 PSAS extensions for interfacing with the NASA-NCAR GCM

The current GEOS-2 version of PSAS analyzes global sea level pressure and near surface winds over the oceans, as well as geopotential height, vector wind, and water vapor mixing ratio on constant pressure surfaces. The upper air height/wind analyses and the sea level pressure/surface wind analyses are multivariate. The moisture analysis is performed using a univariate statistical algorithm at levels from 1000 hPa to 300 hPa. The basic GEOS-2 DAS configuration consists of a 2° latitude by 2.5° longitude, and 18 vertical levels (0.4, 1, 5, 7, 10, 30, 50, 70, 100, 150, 200, 250, 300, 400, 500, 700, 850, 1000 hPa).

The finite-volume formulation of the NASA-NCAR GCM, and the flexibility afforded by the observation-space formulation of PSAS calls for a redesign of the model-analysis interface. Towards this goal, PSAS will produce analysis increments directly on model levels. As a computational device, the analysis increments at some model levels will be computed as an interpolation from adjacent levels; the validity of this approach will be assessed through numerical experimentation. PSAS will produce analysis increments winds, geopotential heights and a moisture variable to be determined (e.g., log of specific humidity or dew point temperature). The computation of analysis increments of surface pressure and other interface related issues are addressed in section 5.

5 Interfacing the NASA-NCAR GCM fields to the statistical analysis

This section introduces the algorithmic details of the analysis-model interface which differs in many respects from the sigma-to-pressure/pressure-to-sigma approach in GEOS-2 DAS (Pfaendtner et al 1994, DAO 1996). We start by defining the NASA-NCAR GCM *state vector* in subsection 5.1, along with its associated finite-volume

mesh. The simulation of observations to be included in the *Violet* Core System is discussed in subsection 5.2. Finally, we describe the computation of analysis increments of surface pressure, and the mapping of PSAS analysis increments to the new vertical grid associated with the *after analysis* surface pressure.

5.1 The NASA-NCAR GCM state vector

The first guess vector w^f provided by the NASA-NCAR GCM will consist of the following variables:

$$w^f = (u \quad v \quad h \quad q \quad \delta p)^T \tag{15}$$

where

u	zonal wind (m/s)
v	meridional wind (m/s)
h	geopotential height (m)
q	specific humidity (g/kg)
δp	pressure-thickness of the finite-volume

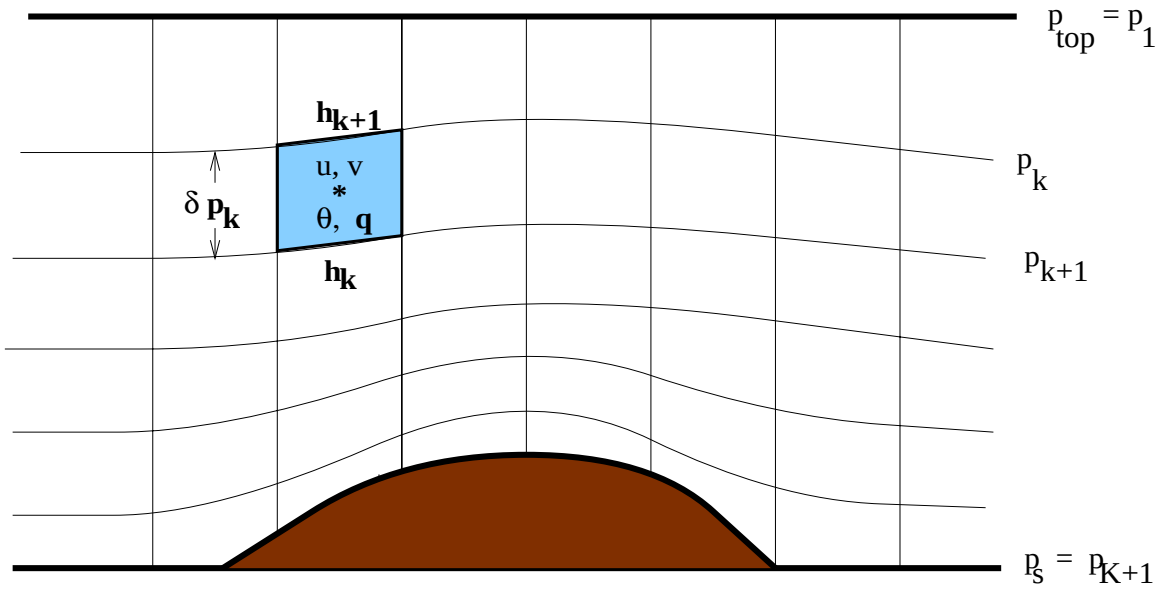

Figure 3: Finite Lagrangian control-volume and state variables.

All these 3D fields will be given on the same regular latitude-longitude grid (the staggered wind components will be averaged to coincide with the other fields). The finite Lagrangian control volume of the NASA-NCAR GCM is illustrated in Fig. 3. There are K vertical layers, with the pressure at the edge of these layers given by $p_1 = p_{top}$, and

$$p_k = p_{top} + \sum_{\ell=1}^{k-1} \delta p_\ell, \quad k = 2, ..., K+1 \tag{16}$$

where the subscript k refers to the vertical level/layer. Notice that the surface pressure is given by $p_s = p_{K+1}$. The pressure thickness δp is a prognostic variable which is evolved by the finite-volume dynamical core. However, after each time-step model fields are mapped from this Lagrangian control-volume vertical coordinate to a *fixed* Eulerian reference coordinate given by:

$$p_k = a_k + b_k p_s, \qquad k = 1, ..., K+1 \tag{17}$$

The coefficients a_k and b_k are chosen in such way to provide for hybrid contant-pressure/terrain-following vertical coordinate. Namely, for $k \leq k_s$ ($p < p_{k_s+1} \equiv p_{int}$) we have constant pressure surfaces ($b_k = 0$ for $k \leq k_s + 1$), while a terrain following vertical coordinate is used for $k > k_s$. At the surface, $p_{K+1} = p_s$, implying $a_{k+1} = 0$ and $b_{K+1} = 1$. Figure 4 illustrates typical values of a_k and b_k for $K = 55$, $p_s = 1000$ hPa and $p_{int} = 100$ hPa; also shown is the normalized pressure coordinate $\eta_k = (p_k - p_{top})/(p_s - p_{top})$.

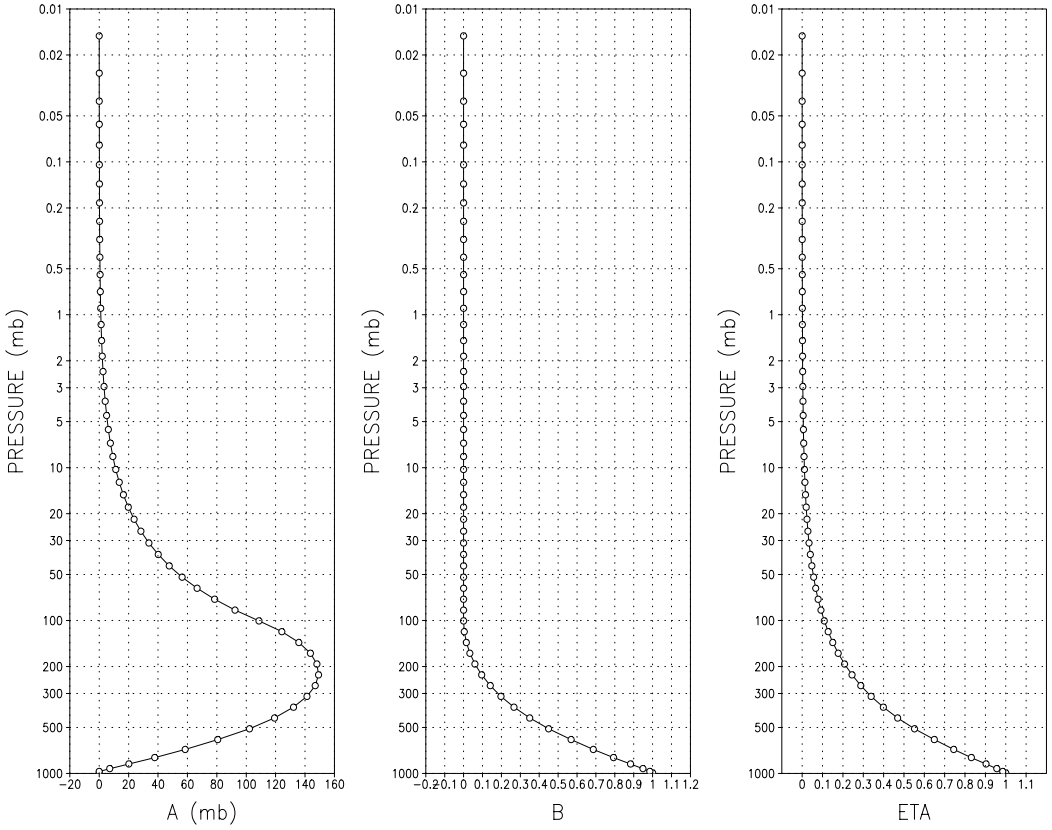

Figure 4: Fixed Eulerian reference coordinate system. See text for details.

For the purpose of data assimilation, we will assume that u, v, q and the (scaled) virtual potential temperature

$$\theta_k = -\frac{g}{C_p} \cdot \frac{h_k - h_{k+1}}{p_k^\kappa - p_{k+1}^\kappa}, \qquad \kappa = R/C_p \tag{18}$$

represent the (mass-averaged) finite-volume mean. For interpolation purposes, we will assume that these volume-mean variables are valid at the *center of mass* of the Lagrangian control volume, which has vertical coordinate

$$p_{k+1/2} = \frac{1}{2}\left(p_k + p_{k+1}\right) \tag{19}$$

The horizontal coordinates of the center of mass will be approximated by averaging the latitudes/longitudes at the edges of the gridboxes. Notice that the geopotential field h is given at the *pressure edges* of the finite control-volume, but horizontally at the center of the gridbox.

5.2 Simulating observations from the NASA-NCAR GCM fields

The observing system

The *Violet* Core System will include the following observing systems:

Land Surface observations. The system will assimilate hourly station pressure observations to take advantage of the Rapid Update Cycle capability. The feasibility of assimilating hourly specific humidity observations at 2 meters will be examined.

Ocean surface observations. The system will be capable of assimilating sea level pressure, winds, and specific humidity from ships, buoys, and sea platforms. Taking proper consideration of anemometer height, these *surface* observations will be homogenized to a standard level (right above the surface layer) using similarity theory.

Rawinsondes. The usual height, winds and moisture at mandatory levels. The assimilation of significant level data will be examined.

TOVS Retrievals. The system will initially assimilate TOVS layer mean temperature retrievals, converted to height retrievals using the first guess for the lower boundary condition. The system shall work with interactive or non-interactive retrievals, whichever is available at the time.

Aircraft winds. Zonal and meridional wind components.

Cloud track winds. Zonal and meridional wind components.

The observation simulator

The key quantity needed by the on-line QC system, and ultimately by PSAS, is the innovation or observation minus forecast (O-F) residual

$$v = w^o - h(w^f). \tag{20}$$

For a generalized observable, say, radiances, the non-linear observation operator h can be represented as an interpolation operator $\mathcal{I}$ applied to the model fields, followed by a radiative transfer model, f, viz.

$$h(w^f) = f(\mathcal{I}w^f). \tag{21}$$

However, the *Violet* Core System will only assimilate data types directly related to the state variables, therefore only the interpolation operator $\mathcal{I}$ will be described.

With the exception of geopotential heights, all the observables listed in subsection 5.2, are point measurements. For these observables, the model fields will be interpolated bi-linearly in latitude-longitude in the horizontal, and linearly in log-p in the vertical. For geopotential height, being a column integral, special attention will be paid to proper weigh partial volumes. Notice that the *error of representativeness* associated with this observing system is simply related to the spatial standard deviation of each observable within the Lagrangian control-volume.

5.3 The after analysis surface pressure

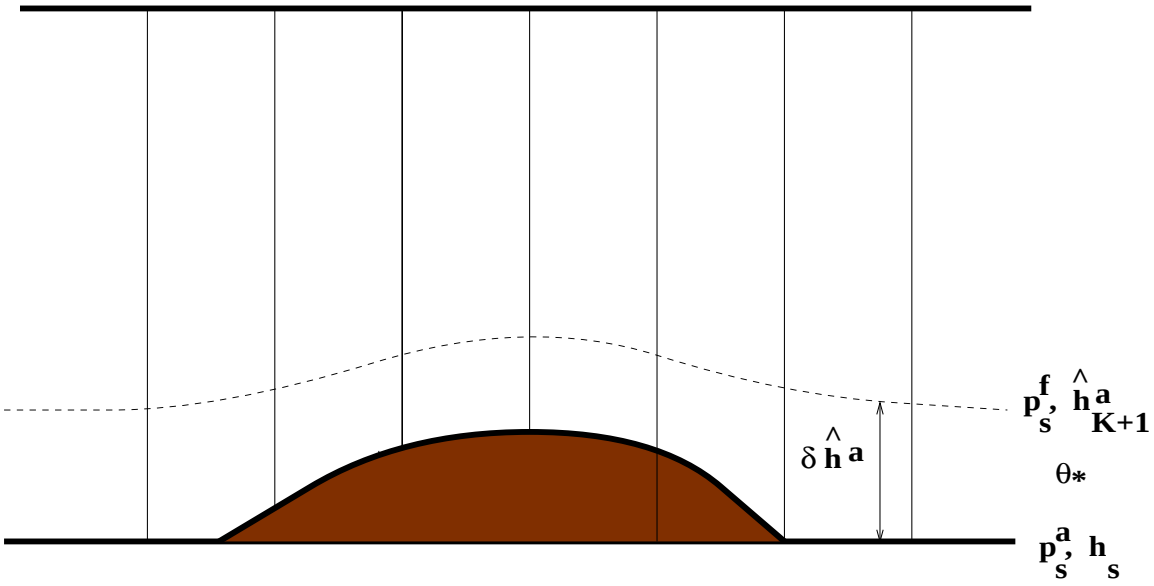

Figure 5: The after-analysis height and surface pressure.

As mentioned in section 4, PSAS will produce analysis increments directly on the model finite-volume grid. That is, on the latitude-longitude of each gridbox, and on the pressure levels/edges implied by the first guess pressure-thickness δp^f. We will denote the analysis increments on this first guess grid by $\delta\hat{w}^a$. In particular, for the lowest atmospheric level we have

$$\hat{h}^a_{K+1} = h_s + \delta\hat{h}^a_{K+1} \tag{22}$$

where h_s is the topographic height. Since in general $\delta\hat{h}^a_{K+1} \neq 0$, we have $\hat{h}^a_{K+1} \neq h_s$ and this pressure level, therefore, no longer corresponds to the surface. The new

surface is located at the pressure level for which $\hat{h}^a = h_s$ (see Fig. 5). From the discrete form of the hydrostatic equation we have

$$\begin{aligned} \delta\hat{h}^a_{K+1} &= \hat{h}^a_{K+1} - h_s \\ &= -\frac{C_p}{g}\theta_* \left[(p^f_s)^\kappa - (p^a_s)^\kappa\right] \end{aligned} \tag{23}$$

where the (yet) undetermined quantity θ_* is the the mean virtual potential temperature in the layer between pressures p^f_s and p^a_s. An expression for the *after-analysis* surface pressure follows from (23):

$$p^a_s = p^f_s \cdot \left[1 + \frac{g\delta\hat{h}^a_{K+1}}{C_p\theta_*(p^f_s)^\kappa}\right]^{1/\kappa} \tag{24}$$

Notice that θ_* enters this equation in the denominator, and therefore the final value of p^a_s is not very sensitive on the precise value of θ_*. For typical values, $p^f_s \sim 1000$ hPa, $\delta\hat{h}^a_{K+1} \sim 100$ m, $T_* = (p^f_s)^\kappa\theta_* \sim 300$ K, we estimate that an error of 5 K in T_* corresponds to less than 0.2 hPa error in p^a_s. For the calculation in (23), thus, we compute θ_* as the layer-mean virtual temperature at the lowest control-volume associated with p^f_s, viz.

$$\theta_* \approx -\frac{g}{C_p} \cdot \frac{\hat{h}^a_{K+1} - \hat{h}^a_K}{(p^f_{K+1})^\kappa - (p^f_K)^\kappa} \tag{25}$$

5.4 Mapping of analysis fields

Once we have determined the after-analysis surface pressure, one is faced with the question of how to come up with the after-analysis pressure-thickness δp^a for each finite control-volume. This quantity is an abstract concept associated with the particular discretization of the NASA-NCAR GCM; it is not clear at this stage whether this quantity could ever be observed. In order to close the problem, we will consider two different approaches:

Method I: Remapping to Eulerian reference coordinates

In this approach, we simply adopt an Eulerian reference coordinate system consistent with this new value of $p_s = p^a_s$, viz.

$$p^a_k = a_k + b_k p^a_s, \qquad k = k_s + 1, ..., K + 1 \tag{26}$$

and compute the after analysis pressure thikness δp^a consistent with it:

$$\delta p^a_k = \delta a_k + \delta b_k p^a_s, \qquad k = k_s + 1, ..., K + 1 \tag{27}$$

where $\delta a_k = a_{k+1} - a_k$, $\delta b_k = b_{k+1} - k_k$. Notice that $p^a_k = p^f_k$ for $k = 1, ..., k_s$ since by construction the top pressure levels (above $p_{int} = p_{k_s+1}$) do not depend on the surface pressure (see Figure 4).

At this point, the first guess w^f and the analysis increments $\delta\hat{w}^a$ returned by PSAS are still on the before-analysis pressure levels, $p_k^f, \quad k = 1, ..., K+1$. The analysis field,

$$\hat{w^a} = w^f + \delta\hat{w}^a, \tag{28}$$

therefore, needs to be mapped to the new vertical grid implied by the after-analysis pressure levels given by (26). For this purpose we adopt the same monotonicity-preserving, and mass-, momentum-, and total energy-conserving mapping algorithm used in the finite-volume dynamical core; for details consult the companiom document *DAO ATBD / Next-Generation Model*. Notice that no mapping is necessary for the pure-pressure levels above $p_{k_s+1} = p_{int}$).

Method II: Shaving of model lowest layers

In this approach, we simply add or remove mass from the lowest layers. When $p_s^a > p_s^f$, we simply add mass to the lowest model layer by setting

$$\delta p_K^a = \delta p_K^f + \left(p_s^a - p_s^f\right) \tag{29}$$

keeping $\delta p_k^a = \delta p_k^f$, for $k = 1, ..., K-1$. The volume-mean values of all quantities are not altered by this expansion of the lowest control volume.

However, when $p_s^a < p_s^f$, (29) can lead to very small or even negative values of δp_K^a, requiring some special handling. In some cases, it is necessary to remove one or more of the model lowest layers in order to accomodate the new value of the surface pressure p_s^a. (In practice, model layers are removed by assiging an extremely small mass to it.)

The main advantage of the *shaving method* is that no mapping or interpolation is necessary except for those one or two lowest model layers affected. However, the resulting Lagrangian control volume could be very different from the fixed Eulerian reference coordinate (17). For this reason, when using this *shaving method* we defer output until after the fields have gone trough the Finite-volume dynamical core and physics modules, which includes a remapping to the fixed Eulerian reference coordinate system, and necessary physical adjustments.

Remarks

1. Notice that in the current GEOS-Terra system, the analysis increments returned by PSAS on mandatory pressure levels are *not* interpolated to the sigma levels implied by the after analysis surface pressure field. This procedure will created artificial analysis increments (on pressure surfaces) in the absence of upper data simply because the first guess field never gets mapped to the new sigma-levels (assuming that surface data was present).

2. Our experience with the GEOS-2 DAS has demonstrated the difficulty of using mixing ratio as the analysis variable. Statistical analysis of O-F residuals indicate the mixing ratio (as well as specific humidity) have distributions which are far from normal. For this reason, it is likely that a new moisture variable will be used for analysis. Candidate variables are the log of specific humidity or the dew point temperature.

3. The NASA-NCAR GCM is carefully designed to conserve the dry-air contribution to the global mean surface pressure, while allowing the moist-air contribution to fluctuate according to the global mean imbalance of evaporation minus precipitation. However, in data assimilation mode, both dry- and moist-air contributions to the global mean surface pressure may fluctuate in time due to the partial coverage of the observing system, as well as observational errors. The time evolution of the global mean surface pressure shall be carefully monitored, and if necessary, a correction will be applied to avoid large fluctuations of the dry-air contribution to the surface pressure.

6 Issues related to balance and initialization

The initial implementation of the *Violet* Core System will not include the Incremental Analysis Updates (IAU) of Bloom et al (1998). Instead, PSAS-based analyses will be performed intermittently every hour or so, in the so called Rapid Update Cycle (RUC).

The current implementation of PSAS employs a relatively traditional error covariance model strategy: horizontal correlations for the mass field and moisture are assumed isotropic, and variances are time mean estimates. These assumptions, although already questionable for a 6 hour assimilation cycle, are even more dubious for a shorter assimilation cycle. Error covariance modeling theory suggests that the 1 hour forecast error is strongly influenced by the analysis error field which is known to be highly anisotropic (due to the inhomogeneity of the observing system) and to have a strong time dependency. Therefore, an 1 hour cycle analysis is expected to be less "optimal" than a 6 hour cycle analysis insofar covariance modeling is concerned. However, the RUC analysis has the benefit of utilizing the observations much closer to the observation time. The relative importance of these two competing effects can only be assessed by implementing a RUC analysis and comparing it with the traditional 6 hour assimilation cycle.

In order to investigate these issues, a RUC with a assimilation interval of 1 hour have been implemented in the current GEOS DAS at 2x2.5 and 4x5 degree resolutions. These initial RUC experiments include IAU (increments are now introduced within a time window, 30 minutes before/after analysis time), although a mathematical analysis of the IAU algorithm (Bloom et al 1998) indicates that IAU looses its low-pass filter capabilities for such a short assimilation cycle. Consistently, examination of surface pressure tendencies produced with the RUC system indicates a slight increase in the noise level for some regions, accompanied by a slight increase in the height O-F standard deviation, and a decrease in wind O-F standard deviation. By sampling the forecast fields much closer to observation time, anomalous O-F variance in the stratosphere associated with missampled tidal motions have been eliminated. Surprisingly, the RUC system has also a positive impact on the climate of the system. The zonal mean meridional circulation has a stronger Hadley circulation, and eliminates some undesirable small scale structure at 70 hPa near the equator. These preliminary results seem to indicate that the benefits of using data closer to observation time outweighs the "violation" of the covariance modeling assumptions. One possible interpretation is that within the current framework of isotropic correlations/time-mean error statistics the basic assumptions are already violated with a 6 hour cycle, and shortening the assimilation cycle does have any significant impact on the system.

As these RUC experiments indicate, the more frequent analysis is slightly noisier, and this increased noise can potentially have a deleterious effect on the extended forecasts.

One way of controlling this spurious noise is by adding an additional constraint to the analysis which penalizes the contribution of the upper level analysis increments to the surface pressure tendency; the goal of this approach is to filter out divergent motion associated with the external mode. Another approach, which is particularly attractive due to its simple implementation, is to perform initialization by means of a *digital filter*. In this algorithm, a short term forecast (say, 6 hours) is run from the non-initialized analysis, and the model state variables are averaged in time according to pre-determined filter coefficients (Fillion et al. 1995, Lynch and Huang 1994). The extended forecast is then restarted at the middle of the initial short term period (say, at 3 hours) from the averaged (filtered) model state. We plan to assess the effect of a digital filter on the forecast skills produced with the *Violet* Core System. If such analysis indicates a large degree of imbalance (as measured by the amount of filtering performed), the implementation of additional penalty terms in PSAS will be evaluated.

7 Testing Strategy

The ultimate goal of data assimilation is to combine the information content of a climate model with observational data to produce the best snapshot of the earth climate system. In this sense, by constraining the prognostic variables, a data assimilation system is expected to add value to a climate simulation. In practice, however, experience has shown that the information content of model/observations are not always additive. Complex non-linear interactions between observations and GCM parameterizations which have been tuned for a climate state other than the observed, can actually result in some climate parameters being degraded in the process of data assimilation (e.g., Molod et al 1996). A point in case is the zonal mean meridional circulation and precipitation fields for the GEOS DAS. The GEOS GCM in AMIP simulation mode produces a stronger Hadley cell and sharper ITCZ than the corresponding GEOS DAS, both of which are believed to be deficiencies in early DAS products. Therefore, prior to assessing the climate parameters in assimilation mode we will perform parallel AMIP style simulations for the same periods (see subsection 7.2) and conduct intercomparisons of the climate parameters in simulation/assimilation modes.

Next we outline the main components of this pre-validation assessement. The main objective of this testing phase is to establish the readiness of the *Violet* Core System for a full validation exercise. Therefore, a detailed validation plan is beyond the scope of this document and will appear elsewhere.

7.1 General system configuration and resolution

Ultimately, the new DAO physical-space/finite-volume DAS will be at least at a 1 degree horizontal resolution. However, in this early developmental stage we will be concentrating on a hybrid system configuration which is compatible with the current GEOS-Tera system, that is, a 1x1.25 degree resolution model coupled to an analysis system producing 2x2.5 degree resolution *analysis increments*. In addition to providing more direct means of comparison with the current system, the faster throughput of this hybrid system can also expedite the tuning process. At the software level, the system will be designed in such a way that the resolution be specified at compilation

time (NASA-NCAR GCM) or at run time (PSAS and QC components).

NASA-NCAR GCM resolution: 1x1.25 horizontal resolution, 55 vertical layers, $p_{top} = 0.01$ hPa.

Analysis resolution: 2x2.5 horizontal resolution, 30 analysis layers with 25 *shadow layers*,
$p_{top} = 0.01$ hPa. Recall that the analysis increment at shadow layers are obtained through vertical interpolation. This interpolation will be performed outside PSAS in a GCM consistent fashion.

Remark

Notice that the analysis field, $w^a = w^f + \delta w^a$ can preserve the 1 degree features implied by the first guess w^f; only the corrections to the first guess, δw^a, will be performed at lower resolution and interpolated to the finer resolution grid. The calculation of innovatons, $v = w^o - Hw^f$, will also be performed at full model resolution. This *incremental* formulation of the data assimilation algorithm is commonly used in operational implementation of 3D-VAR/4D-VAR algorithms (e.g., Coutier et al. 1994, Laroche et al. 1999). The scientific basis of this approach comes from the smoothness of the analysis increments δw^a which is a consequence the large error decorrelation length scales used by these analysis systems. As the error covariance modeling evolves in PSAS, *e.g.,*, through the *Parameterized Kalman Filter* effort (S. Cohn, personal communication), it is expected that ultimately both model and analysis will be operating at the same resolution. Therefore, it is crucial that a resolution-independent system be designed.

7.2 Testing Periods

The validation exercise will focus on the following periods:

Winter 1998: from 11/15/1997 to 2/28/1998. This is a standard season for recent validation. OLR estimates from CERES can be used for assessing the radiation fields.

Summer 1998: from 5/15/1998 to 8/31/1998. Again, there several pre GEOS-Terra DAS experiments for comparison.

Fall 1999/Winter 2000: from 9/1/1999 to present. This period coincides with the beginning of operational GEOS-Terra activities.

7.3 Assessing the Observing system

This component focus on comparing the system output directly to observations. Traditionally, one has focused on observation minus forecast residuals (O-F) from the DAS. However, the same tools can be used to assess the output from a pure model simulation, although this is optional for this project.

O-F residuals. Regional averaged bias and standard deviations, as well as horizontal maps for selected levels in the PBL, troposphere and stratosphere. **Metric:** O-F mean and standard deviations should be smaller or comparable to O-F statistics from the GEOS-Terra system.

O-A residuals. Analysis similar to O-F. The goal here is to verify that the observational data is represented in the analysis. **Metric:** none, as drawing closer to the data does not necessarily translate into better analysis. The O-A tests of Hollingsworth and Lonnberg (1989) are not required at this stage.

7.4 Assessing analysis and forecasts

This component applies only to DAS fields.

Analysis intercomparison. Check analysis of height, winds and moisture against corresponding ECMWF (or NCEP) reanalysis. **Metric:** none, as the ECMWF/NCEP analysis cannot be taken as ground truth.

Forecast skills. Anomaly correlations and RMS for Northern/Southern hemisphere extratropics, and tropics, verifying against own analysis (or ECMWF analysis). Also, verify forecast against observations (O-xF). **Metric:** anomaly correlations should be higher, and RMS smaller, compared to GEOS-Terra.

Balance issues. Plot spatial RMS of surface pressure tendency as a function of time, and temporal RMS as a function of longitude/latitude. **Metric:** the RMS of surface pressure tendency should be similar or smaller compared to GEOS-Terra.

Transport characteristics. Use assimilation winds to drive ozone assimilation system and examine impact on ozone O-F's. **Metric:** ozone system driven by the *Violet* winds should produce smaller or comparable O-F to control system forced by GEOS-Terra winds.

7.5 Assessing the Climate Diagnostics

This component applies to both simulation and assimilation monthly mean fields.

Zonal mean circulation. Compare zonally averaged monthly mean winds, temperature and moisture to ERA/NCEP reanalysis. **Metric:** simulations should be closer to reanalyses; assimilation fields not necessarily.

Stationary waves. Compare stationary waves of temperature, streamfunction, velocity potential, and moisture at 850 hPa, 300 hPa and 1 hPa, zonal cross-sections at 45S, equator and 45N, meridional cross-sections at 90W, at the prime meridian, and 90E. **Metric:** simulations should be closer to reanalyses; assimilation fields not necessarily.

Precipitation. Compare monthly mean precipitation to GPCP/NCEP raingauge estimates, assessing mean diurnal cycle whenever possible. **Metric:** Patterns

of both simulation and assimilation fields should be closer to independent estimates compared to GEOS-Terra; precipitation amounts should be evaluated within the accuracy of each estimate.

Clouds. Compare monthly mean low, middle and high clouds to ISCCP (1992 only). **Metric:** both simulation and assimilation fields should produce estimates closer to independent estimates compared to the GEOS-Terra system.

Surface radiation budget. Compare monthly surface radiation parameters to SRB and Man-Li Wu estimates; assess skin temperature as well. **Metric:** both simulation and assimilation fields should produce estimates closer to independent estimates compared to the GEOS-Terra system.

Top-of-atmosphere radiation budget. Compare radiation diagnostics to NOAA/OLR (1992/98) and CERES estimates (1998 only). **Metric:** both simulation and assimilation fields should produce estimates closer to independent estimates compared to the GEOS-Terra system.

8 Open problems and issues not addressed by the "Violet" System

The *Violet* Core system described in this document does not address all the issues necessary for the full implementation of an operational data assimilation system. Several of the issues which will require attention after this initial "fast prototype" phase are listed below.

1. The *Violet* Core System will be developed under the "shared memory" paradigm, although the overall system architecture will be carefully designed to accommodate a subsequent MPI implementation. This computational issue is beyond of the scope of this document.

2. The *Violet* Core System will *not* produce output compliant with the DAO File specs:

   Analysis output: the analysis output stream, including ODS, and before- and after-analysis states using GFIO (HDF/COARDS compliant) will be similar to GEOS-Terra. However, the gridded fields will be on the model grid, not on mandatory levels.

   NASA-NCAR GCM output: there are no plans to reconcile the GCM output with the file specs for the *Violet* system. However, the output files are in IEEE or NetCDF format which can be directly read by analysis tools such as GrADS and Matlab, without presenting great difficulties for the validation exercise.

3. As GEOS-Terra, the version of PSAS using in the *Violet* Core System still does not support explicit observation operators. Therefore, only those observables directly related to state variables can be assimilated.

4. The following GEOS-Terra observing systems are not being included in this initial implementation:

   - Satellite derived surface wind speeds: SSM/I

- SSM/I total precipitable water

These observing systems are being excluded in the interest of time, as they require some degree of exception handling. There is no built in limitation preventing the assimilation of these data types. For the Fall 1999, experimental wind velocity retrievals from QuickScat will be used to assess the impact of satellite wind observations.

5. No explicit analysis bias correction will be initially implemented (Dee and da Silva 1998).

6. As in GEOS-Terra, this system still relies on time mean forecast error variances adaptively estimated with ATESS. The overall system architecture, however, is being designed in such a way as to replace the ATESS estimates with flow dependent statistics afforded by the *Parameterized Kalman Filter* effort (S. Cohn, personal communication).

7. The impact of not having IAU on model diagnostics is still to be determined. Early experience with RUC on GEOS-DAS seems to suggest the feasibility of this approach.

9 Concluding Remarks

We have described the main scientific attributes of the new DAO Physical-space/Finite-volume Core Data Assimilation system, focusing on an initial configuration which is intended to serve as a proof of concept for this new system. As of this writing (January 2000), the main software software components have been written and unit tested. Integration test is currently in progress.

References

Bloom, S.S., L.L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data Assimilation Using Incremental Analysis Updates. *Mon. Wea. Rev.*, **124**, 1256–1271.

Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan*, **75**, 257–288.

Cohn, S. E., and R. Todling, 1996: Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteor. Soc. Japan*, **74**, 63–75.

Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz, D. Lamich. 1998: Assessing the Effects of Data Selection with the DAO Physical-space Statistical Analysis System. *Mon. Wea. Rev.* . **126**, 2913-26.

DAO, 1996: Algorithm Theoretical Basis Docoment for Goddard Earth Observing System Data Assimilation System (GEOS DAS) With a Focus on Version 2. Available on-line from `http://dao.gsfc.nasa.gov/subpages/atbd.html`.

DAO, 1997: GEOS-3 Data Assimilation System Architectural Design. *DAO Office Note 97-06*. Data Assimilation Office, Goddard Space Flight Center, Greenbelt, MD 20771.

Dee, D. and A. da Silva, 1998: Data assimilation in the presence of forecast bias. *Q. J. R. Meteor. Soc.*, **124**, 269-295.

Dee, D., and A. da Silva, 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Mon. Wea. Rev.*. In press.

Dee, D., G. Gaspari, C. Redder, L. Rukhovets, and A. da Silva, 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part II: Applications. *Mon. Wea. Rev.*, In press.

Fillion, L., H. L. Mitchell, H. Ritchie, and A. Staniforth, 1995: The impact of a digital filter finalization technique in a global data assimilation system. *Tellus*, **47A**, 304-323.

Guo, J., and A. da Silva, 1997: Computational aspects of Goddard's Physical-space Statispical Analysis System (PSAS). In *Numerical simulations in the environmental and earth sciences.*, Garcia et al., Eds., ISBN 052158047, Cambridge University Press, 1997.

Guo, J,, J. W. Larson, G, Gaspari, A. da Silva and P. M. Lyster, 1998: Documentation of the Physical-space Statistical Analysis System (PSAS) Part II: The factored-operator error covariance model formulation. *DAO Office Note 1998-04*.

Laroche, S., P. Gauthier, J. St-James and J, Morneau, 1999: Implementation of a 3D variational data assimilation system at the Canadian Meteorological Center. Part II: the regional analysis. Submmitted to *Tellus*.

Larson, J. W., J. Guo, G. Gaspari, A. da Silva and P. M. Lyster, 1998: Documentation of the Physical-space Statistical Analysis System (PSAS) Part III: The software implementation of PSAS. *DAO Office Note 1998-05*.

Hack, J. J., B. A. Boville, J. T. Kiehl, P. J Rasch and D. L. Williamson, 1993: Description of the NCAR Community Climate Model (CCM2). *NCAR Technical Note*, NCAR/TN-382+STR, Boulder, CO, 108 pp.

Hollingsworth, A. and P. Lonnberg, 1989: The verification of objective analyses: diagnostics of analysis system performance, *Meteorol. Atmos. Phys.*, **40**, 3-27.

Kiehl, J. T., J. J. Hack, G. B. Bonan, B. A. Boville, B. P. Briegleb, D. L. Williamson, and P. J. Rasch 1996: Description of the NCAR Community Climate Model (CCM3). *NCAR Technical Note*, NCAR/TN-420+STR, Boulder, CO, 152pp.

Lin, S.-J., 1997: A finite-volume integration method for computing pressure gradient forces in general vertical coordinates. *Q. J. Roy. Met. Soc.*, **123**, 1749-1762.

Lin, S.-J., and R. B. Rood, 1997: An explicit flux-form semi-Lagrangian shallow water model on the sphere. Q. J. Roy. Met. Soc., 123, 2477-2498.

Lin, S.-J., and R. B. Rood, 1996: Multidimensional Flux Form Semi-Lagrangian Transport schemes. *Mon. Wea. Rev.*, **124**, 2046-2070.

Lin, S.-J., and R. B. Rood, 1998: A flux-form semi-Lagrangian general circulation model with a Lagrangian control-volume vertical coordinate. *The Rossby-100 symposium*, Stockholm, Sweden.

Lynch, P. and X.-Y. Huang, 1994: Diabatic initialization using recursive filters. *Tellus*, **46A**, 304-323.

Molod, Andrea, H.M. Helfand, and L.L. Takacs, 1996: The Climatology of Parameterized Physical Processes in the GEOS-1 GCM and their Impact on the GEOS-1 Data Assimilation System. *J. Climate*, **9**, 764-785.

Pfaendtner, J., S. Bloom, D. Lamich, M. Seablom, M. Sienkiewicz, J. Stobie, and A. da Silva, 1995: Documentation of the Goddard Earth Observing System (GEOS) Data Assimilation System - Version 4. *NASA Tech. Memo. 104606*, Vol. 4.

da Silva, A. M., C. C. Young and S. Levitus, 1994: Atlas of Surface Marine Data 1994, Volume 1: Algorithms and Procedures. *NOAA Atlas NESDIS 6*, U.S. Department of Commerce, NOAA, NESDIS, 83pp.

da Silva, A. da J. Guo, 1996: Documentation of the Physical-space Statistical Analysis System (PSAS) Part I: The Conjugate Gradient Solver Version PSAS-1.00. *DAO Office Note 96-02.*

da Silva, A., M. Tippett and J. Guo, 1999: The PSAS User's Manual. *DAO Office Note 1999-xx.*

da Silva, A. and C. Redder, 1995: Documentation of the GEOS/DAS Observation Data Stream (ODS) Version 1.0. *DAO Office Note 95-01.*. Data Assimilation Office, Goddard Space Flight Center, Greenbelt, MD 20771.

List of Acronyms

ATBD	Algorithm Theoretical basis Document
AMIP	Atmospheric Model Intercomparison Project
ATESS	Adaptive Tuning of Error Statistics Subsystem
CCM	Climate Community Model
CCM3	Climate Community Model Version 3
CERES	Clouds and Earth's Radiation Energy System
COARDS	Cooperative Ocean-atmosphere Data Service
DAO	Data Assimilation Office
DAS	Data Assimilation System
ECMWF	European Center for Midrange Weather forecasts
EKF	Extended kalman Filter
GCM	General circulation Model
GEOS	Goddard Earth Observing System
GPCP	Global Precipitation Climatology Project
HDF	Hierarchical Data Format
IAU	Incremental Analysis Updates
ITCZ	Intertropical Convergence Zone
NCEP	National Centers for Environmental Predictions
NOAA	National Oceanographic and Atmospheric Administration
OI	Optimal Interpolation
OLR	Outgoing Longwave radiation
PSAS	Physical-space Statistical Analysis System
PKF	Parameterized Kalman Filter
QC	Quality Control
RUC	Rapid update Cycle
SQC	Statistical Quality Control
TOVS	TIROS Operational Vertical Sounder